

Korrespondenzanalyse: ein multivariates Verfahren zur Analyse qualitativer Daten

Blasius, Jörg

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Blasius, J. (1987). Korrespondenzanalyse: ein multivariates Verfahren zur Analyse qualitativer Daten. *Historical Social Research*, 12(2/3), 172-189. <https://doi.org/10.12759/hsr.12.1987.2/3.172-189>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>

KORRESPONDENZANALYSE - EIN MULTIVARIATES VERFAHREN ZUR ANALYSE QUALITATIVER DATEN

Jörg Blasius (*)

Abstract: The purpose of the article is to introduce the method of correspondence analysis, using data from the Reichsstudentenwerk, collected from 1933 to 1942 and published by Arminger in 1984. Correspondence analysis is a multivariate technique which main characteristics is to display the columns and rows of twodimensional contingency tables graphically. Apart from the graphical display, further dimensions may be interpreted numerically. With this method we want to describe groups in six first memberships in organisations (NSDAP, SS, SA, NSDStB, others, no member) by age, their fathers home country and status, sex, subject of study, and period of first loan. Correspondence analysis provides an adequate method for analyzing and describing the six subpopulations.

1. Allgemeines

Häufig werden in den Sozialwissenschaften - und hier speziell auch in der Historischen Sozialforschung - Verfahren benötigt, in denen es keine Voraussetzungen bezüglich des Skalenniveaus gibt und wo gleichzeitig eine multivariate Anwendung möglich ist. Eines dieser Verfahren ist die in Frankreich (Benzécri: 1963, Escofier-Cordier: 1965) entwickelte Korrespondenzanalyse. Ihre ersten Ansätze können auf Hirschfeld (1935) zurückgeführt werden, der als erster eine algebraische Formulierung der "Korrelation" von Zeilen und Spalten einer Kontingenztafel gab. Guttman (1941, 1950) entwickelte dies als Skalierungsverfahren in seiner "principal components analysis" weiter, ebenso wie Hayashi (1950, 1952, 1954) mit seiner "quantification of qualitative data".

Ähnlich der Korrespondenzanalyse sind die Verfahren des "reciprocal averaging" (insb. Hill: 1973) und des "dual or optimal scaling" (insb. Nishisato: 1978, 1980). Die zur Zeit ausführlichste Beschreibung der Korrespondenzanalyse und die statistischen Grundlagen liefert die Arbeit von Greenacre (1984). Sie hat aber den Nachteil, für einen Nichtmathematiker nur schwer verständlich zu sein.

Bei der Korrespondenzanalyse handelt es sich primär um ein Verfahren zur graphischen Darstellung der Spalten und Zeilen von zweidimensionalen (two-way) Kontingenztabellen (Greenacre 1981, 119). Mit Zweidimensionalität von Kontingenztabellen ist hier gemeint, daß alle Kreuztabellen der "abhängigen" mit den "unabhängigen" (abhängig und unabhängig werden hier nicht im kausalen Sinne verwendet, sondern meinen lediglich Spalten- und Zeilenvariablen zwischen denen ein beliebiger Zusammenhang besteht) Variablen untereinander geschrieben werden. Dies hat den Vorteil, daß es bei der multivariaten Anwendung keine Probleme mit der Frequenz der Zellenbesetzung gibt, wie dies z.B. bei der log-linearen Analyse der Fall ist.

(*) Address all communications to: Jörg Blasius, Universität zu Köln, Seminar für Soziologie, Albertus-Magnus-Platz, D-5000 Köln 41.

Wie bei der Hauptkomponentenanalyse gibt es einen Satz von orthogonal aufeinander stehenden Vektoren, die einen niederdimensionalen Raum aufspannen, im Fall einer beabsichtigten grafischen Darstellung genau zwei. (Es lassen sich prinzipiell auch höherdimensionale Räume einführen, welche in Analogie zu grafischen Rotationsverfahren der Faktorenanalyse nur ebenenweise ausgeplottet werden können) Neu, im Sinne der praktischen Anwendung, ist die datenabhängige Metrik. Hier unterscheidet sich die Korrespondenzanalyse von Verfahren wie multidimensionale Skalierung oder Clusteranalyse.

Mit der Korrespondenzanalyse lassen sich sowohl Daten auf Individualebene als auch auf Aggregatebene auswerten, wobei in dieser Arbeit nur letzteres vorgestellt werden soll. (Ein Anwendungsbeispiel mit Daten auf Individualdatenebene ist z.B. die Arbeit von Dangschat und Blasius (1987)) Wir verwenden dieses Verfahren für eine Beschreibung der ersten Mitgliedschaft in einer Organisation der NSDAP (NSDAP, SS, SA, Nationalsozialistischer Studentenbund (NSDStB), andere Mitgliedschaft, keine Mitgliedschaft) in der Zeit von 1933 bis 1942. Datengrundlage sind Aufzeichnungen des Reichsstudentenwerkes über Studenten, die in der genannten Zeit Darlehen erhielten. Von diesem Ausgangsmaterial veröffentlichte Arminger (1984) die Auswertung einer Zufallsstichprobe (N=566), wobei er im multivariaten Teil log-lineare Analysen verwendete. Unser Ausgangsmaterial sind die zweidimensionalen Kontingenztafeln, die Arminger in seinem beschreibenden Teil vorstellt. Zur Beschreibung der Variable "Organisation der ersten Mitgliedschaft" verwenden wir folgende Variablen: Studienfach (Theologie - evangelisch und katholisch/ Staatswissenschaften/ Medizin und Pharmazie/ Naturwissenschaften/ Geisteswissenschaften - Sprachen, Philosophie, Geschichte, .../ Architektur/ Ingenieurwesen/ Sonstiges), Geschlecht, Erhalt des ersten Darlehens (1933-1938/ 1939-1942), Herkunft des Vaters (Deutscher/ Nicht-Deutscher), Geburtsjahr (bis 1904/ 1905 bis 1909/ 1910- 1914/ 1915 bis 1925), Sozialer Status des Vaters. (Oberschicht - Adlige, Industrielle, Topmanager, .../ Staatsangestellte mit Hochschulabschluß - Priester, Professoren, Offiziere, .../ Sonstige mit Hochschulabschluß - Ärzte, Journalisten, .../ alte Mittelschicht - Farmer, Handwerker, .../ neue Mittelschicht - einfache Beamte, Grundschullehrer, .../ Arbeiterklasse).

Zur formalen Beschreibung des Verfahrens beschränken wir uns vorerst auf die Variable "erste Mitgliedschaft in einer Organisation der NSDAP" (zur besseren Übersicht beschränken wir uns auf die vier Gruppen "NSDAP, SS, SA, keine Mitgliedschaft") in Abhängigkeit der Variable "Studienfach" (zur besseren Übersicht ohne "Sonstige"). Die Berechnungen dieser Beispielsaufgabe wurden mit einem von Blasius/Rohlinger (1987) geschriebenen Programm (in der Matrix-orientierten Sprache SAS, PROC MATRIX) durchgeführt. Dieses Programm hat den Vorteil, daß alle relevanten Zwischenschritte zur Problemlösung ausgegeben werden. Die als erstes ausgegebene Tabelle ist die Matrix der Eingabedaten.

AUSGANGSMATRIX (EINGABEDATEN)

ZEILENSUMMEN

A	NSDAP	SS	SA	KMTG	CL	COLI
THELOG	2	0	9	7	THELOG	18
STATSW	22	14	39	20	STATSW	95
MED-PH	15	23	58	22	MED-PH	118
NATURW	6	4	21	14	NATURW	45
GEISTW	7	8	18	8	GEISTW	41
ARCHIT	4	6	16	13	ARCHIT	39
INGENE	6	4	16	4	INGENE	30
SPALTENSUMMEN					GESAMTSUMME	
B1	NSDAP	SS	SA	KMTG	N	COLI
ROW1	62	59	177	88	ROW1	386

Quelle: Arminger 1934, S. 25.

2. Einführung in die Korrespondenzanalyse

Wir gehen von einem beliebigen Punkt aus, der sich irgendwo in einem real vorhandenen dreidimensionalen Raum befindet. Dieser Punkt kann anhand seiner Koordinaten (Länge, Breite und Höhe) genau lokalisiert und seine Entfernung zu anderen, sich ebenfalls im Raum befindlichen Punkten bestimmt werden. Wir können diesen Punkt, ebenso wie alle anderen Punkte, die sich in demselben Raum befinden mögen, in vektorieller Schreibweise darstellen.

Werden alle diese Punkte in eine Matrix geschrieben, so hat diese drei Spalten, die Anzahl der Zeilen entspricht der Anzahl der Punkte. Dies entspricht dem formalen Aufbau der Matrix der Eingabedaten (A). Der Unterschied besteht lediglich darin, daß es sich hierbei nicht um "Längenvektoren", sondern um "Studienfachvektoren" handelt, deren Komponenten entsprechend der Häufigkeit des jeweiligen Studienfaches determiniert sind. Dies entspricht dem Eingangsmaterial der Korrespondenzanalyse: Einer zweidimensionalen Kontingenztafel mit Zeilenvektoren (hier die sieben verschiedenen Arten des Studienfaches - also "Studienfachvektoren") und Spaltenvektoren (hier die vier verschiedenen Organisationszugehörigkeiten in der Zeit von 1933 bis 1942 - also "Organisationszugehörigkeitsvektoren").

Um zu einer grafischen Darstellung zu gelangen, sollen die Punkte oder vielmehr die Ausprägungen des Studienfaches derartig in einen zweidimensionalen Unterraum projiziert werden, daß die quadrierten Abstände (Distanzen) und damit auch der Informationsverlust minimal ist. Bei dieser Projektion müssen die relativen Anteile der einzelnen Variablenausprägungen (von Zeilen und Spalten) berücksichtigt werden. So befinden sich z.B. 30,6 Prozent Studenten der medizinischen Fächer, aber nur 4,7 Prozent der Theologiestudenten im Sample der zu beschreibenden Organisationen. Diese sollen entsprechend ihrer wahren Anteile (benannt als Massen) in die Auswertung eingehen und nicht in einem Verhältnis von eins zu eins. Dies wäre aber der Fall, wenn die Randsummen der Kontingenztafeln (sie gehören nicht zu unserer Matrix der Ausgangsdaten) unberücksichtigt blieben.

Demnach besteht das Dilemma, daß die Punkte, die in den neuen Raum projiziert werden sollen, unterschiedliche Massen haben, oder anders ausgedrückt, unterschiedlich schwer sind. Um dieses Problem zu lösen, benötigen wir eine Hilfsgröße. Dies soll der Schwerpunkt sein, den wir als den Punkt im Raum definieren, an welchem alle Punkte, die in die Analyse eingehen sollen, im Gleichgewicht sind. Eine anschauliche Analogie dazu bietet sich bei einer Apotheke, wo die eine Seite durch Gewichte an der anderen Seite aufgewogen wird. Was bei dieser Waage der Auflagepunkt ist, ist hier der Schwerpunkt, also der Punkt an dem die Summe der linksdrehenden Elemente gleich der Summe der rechtsdrehenden Elemente ist.

In einem ersten Schritt werden alle Zeilen auf die gleiche Länge normiert, die 1 sein soll. Die Berechnung der neuen Werte geschieht durch einfache Prozentuierung. Da die unterschiedlichen Massen von Zeilen (unterschiedliche Anteile von Studenten bei den einzelnen Studienfächern) und der Spalten (unterschiedliche Anzahl von Mitgliedern bei den Organisationen) berücksichtigt werden müssen, werden auch die Randsummen jeweils ins Verhältnis zur Gesamtsumme gesetzt. Das Ergebnis dieses ersten Schrittes zur Problemlösung ist in der Matrix der Zeilenprozentage (PROZ) dargestellt.

ZEILENPROZENTE

PROZ	NSDAP	SS	SA	KMTG
THELOG	0.111111	0	0.5	0.388889
STATSW	0.231579	0.147368	0.410526	0.210526
MED-PH	0.127119	0.194915	0.491525	0.186441
NATURW	0.133333	0.088889	0.466667	0.311111
GEISTW	0.170732	0.195122	0.439024	0.195122
ARCHIT	0.102564	0.153846	0.410256	0.333333
INGENE	0.2	0.133333	0.533333	0.133333

ZEILENSUMMEN IN PROZENT

RPROZ	COLI
THELOG	0.0466321
STATSW	0.246114
MED-PH	0.305699
NATURW	0.11658
GEISTW	0.106218
ARCHIT	0.101036
INGENE	0.0777202

SPALTENSUMMEN IN PROZENT

RPROS	NSDAP	SS	SA	KMTG
ROW1	0.160622	0.15285	0.458549	0.227979

Die relativen Anteile der Zeilen (Zeilensummen in Prozent) werden als Masse des Profils bezeichnet, hier sind die Informationen der Zeilengewichte enthalten. Die relativen Anteile der Spalten (Spaltensummen in Prozent), die ebenso die durchschnittlichen Anteile der Zeilen sind, werden als Durchschnittszeilenprofil bezeichnet. Letzteren Punkt (dargestellt durch einen Vektor) hätten wir auch direkt aus den einzelnen Spaltenvektoren, multipliziert mit dem Vektor der Masse des Profils (RPROZ) (die Länge bliebe unberücksichtigt, da in allen Fällen gleich 1) bestimmen können, er ist der Schwerpunkt (So ergibt sich der erste Wert des Zeilenvektors RPROS: $0,111 \cdot 0,047 + 0,232 \cdot 0,246 + \dots + 0,2 \cdot 0,078 = 0,161$).

Entsprechend der Matrix der Zeilenprozente (PROZ) summieren sich die Komponenten von jedem Zeilenvektor zu Eins. Dies hat zur Folge, daß die sieben Punkte innerhalb des vierdimensionalen Raumes einen höchstens dreidimensionalen Raum aufspannen, der orthogonal zum Vektor (1,1,1,1) ist.

Das Problem besteht jetzt darin, diesen dreidimensionalen Raum in einen neuen Raum zu projektieren, indem die "Massen" (von Zeilen und Spalten) berücksichtigt sind. Gesucht ist somit eine Transformation, die unsere Zeilenvektoren in einen neuen (vorerst wiederum dreidimensionalen) Raum überführt, in dem die euklidische Metrik angenommen wird. In einem ersten Schritt werden lediglich die relativen Spaltenanteile berücksichtigt. Als Lösung dieses Problems läßt sich folgende Transformation angeben (vgl. Greenacre: 1984):

$$(1) \quad \tilde{a}_i^* = D_z^{-1/2} \cdot a_i$$

Hierbei sind die \tilde{a}_i die "Studienfachvektoren" im alten ursprünglichen Raum, die \tilde{a}_i^* die "Studienfachvektoren" im neuen Raum, und ist D_z eine Matrix, in welcher in der Hauptdiagonalen die Elemente des Schwerpunktes stehen, während der Rest aus Nullen besteht. Hieraus läßt sich die benötigte Matrix $D_z^{-1/2}$ bestimmen. Dies ist durch die nachfolgenden drei Matrixen (DZC, IDZ und SQIDZ) dokumentiert.

DIAGONALMATRIX DER SPALTENSUMMEN

DZC	COL1	COL2	COL3	COL4
ROW1	0.160622	0	0	0
ROW2	0	0.15285	0	0
ROW3	0	0	0.458549	0
ROW4	0	0	0	0.227979

INVERSE DER DIAGONALMATRIX

IDZ	COL1	COL2	COL3	COL4
ROW1	6.22581	0	0	0
ROW2	0	6.54237	0	0
ROW3	0	0	2.18079	0
ROW4	0	0	0	4.38636

WURZEL AUS DER INVERSEN DER DIAGONALMATRIX

SQIDZ	COL1	COL2	COL3	COL4
ROW1	2.49516	0	0	0
ROW2	0	2.55781	0	0
ROW3	0	0	1.47675	0
ROW4	0	0	0	2.09436

Führen wir diese Transformation für die sieben Zeilenvektoren durch, so erhalten wir sieben neue Vektoren in einem neuen Raum unter Beibehaltung der "Masse des Profils" - also der Massen der Zeilen. Mit genau der gleichen Transformation wird auch der Schwerpunktsvektor in den neuen Raum überführt, der wieder Schwerpunkt der transformierten Punktwolke ist. Dies soll ausführlich dargestellt werden. So ergibt sich die Matrix AI (Zeilendarstellung - gewichtete Matrix) aus dem Produkt von PROZ * SQIDZ (Zeilenprozente * Wurzel aus der Inversen der Diagonalmatrix) und der Schwerpunktsvektor (AIS) aus RPROS * SQIDZ. Schreiben wir diese Operation vollständig für das erste Element der Schwerpunktsvektors AIS (Schwerpunkt der Zeilendarstellung), so ist $0.161 \cdot 2.495 + 0.153 \cdot 0 + 0.459 \cdot 0 + 0.228 \cdot 0 = 0.401$.

ZEILENDARSTELLUNG- GEWICHTETE MATRIX

AI	COL1	COL2	COL3	COL4
ROW1	0.27724	0	0.738375	0.814475
ROW2	0.577826	0.37694	0.606245	0.440919
ROW3	0.317181	0.498555	0.72586	0.390475
ROW4	0.332688	0.227361	0.68915	0.65158
ROW5	0.426002	0.499084	0.648329	0.408657
ROW6	0.255913	0.393509	0.605846	0.698122
ROW7	0.499031	0.341041	0.7876	0.279249

SCHWERPUNKT DER ZEILENDARSTELLUNG

AIS	COL1	COL2	COL3	COL4
ROW1	0.400776	0.39096	0.677163	0.477472

Die quadrierte Distanz zwischen zwei Vektoren (hier "Studienfachvektoren") ist definiert als:

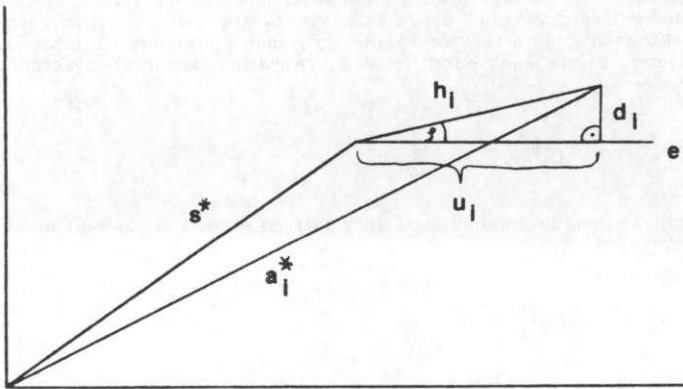
$$(2) d^2(a_i, a_j) = (a_i - a_j)^T D_z^{-1} (a_i - a_j)$$

Durch die Diagonalmatrix D_z^{-1} (SQIDZ) werden Unterschiede, die sich allein durch einen unterschiedlichen Anteil der Spalten (hier Organisationszugehörigkeit) ergeben können, relativiert (vgl. Greenacre 1981:125). Drücken wir die Distanzen mit den transformierten Zeilenvektoren aus, so erhalten wir:

$$(3) \quad d^{2*}(a_i^*, a_j^*) = (a_i^* - a_j^*)^T (a_i^* - a_j^*)$$

Dies entspricht der euklidischen Metrik. Bevor wir die neuen Vektoren in einem niederdimensionalen Raum projektieren dürfen, müssen noch die relativen Anteile der Zeilen (Zellensummen in Prozent) berücksichtigt werden. Die Lösung wird nach Veranschaulichung der Projektion relativ einfach sein. Um den zur grafischen Darstellung gewünschten zweidimensionalen Unterraum aufspannen zu können, müssen wir zuvor, vom Schwerpunkt ausgehend, einen Satz paarweise orthogonal aufeinanderstellender Einheitsvektoren der Länge 1 bestimmen.

Abbildung 1: Grafische Darstellung der Projektion



Das Problem besteht darin, die d_i^2 zu minimieren, d.h. die Abstände zu den neuen Unterraum (der i.d.F. durch zwei Einheitsvektoren aufgespannt wird) sollen möglichst klein sein. Bekannt ist h_i als Differenz von $(a_i^* - s^*)$; aus den bekannten Informationen läßt sich auch der Einheitsvektor e (s. Abbildung) berechnen, worauf an späterer Stelle noch ausführlicher eingegangen wird. Der Vektor u_i entspricht der Entfernung auf dem Einheitsvektor zum Schwerpunkt, der sich durch die Projektion ergibt.

Aus der Definition des Skalarproduktes ergibt sich für

$$(4) \quad u_i = (a_i^* - s^*)^T e, \text{ quadriert}$$

$$(5) \quad u_i^2 = e^T (a_i^* - s^*) (a_i^* - s^*)^T e$$

Des weiteren ergibt sich aus dem Satz des Phytagoras, daß

$$(6) \quad h_i^2 = u_i^2 + d_i^2, \text{ oder durch Summation:}$$

$$(7) \quad \sum h_i^2 = \sum u_i^2 + \sum d_i^2$$

Dies ist gleichbedeutend mit:

$$(8) \quad \sum d_i^2 = \sum h_i^2 - \sum u_i^2$$

Die Summe der d_i^2 soll minimiert werden, dies ist erfüllt, wenn wir die Summe der u_i^2 maximieren.

Kehren wir zurück zur Berechnung der Einheitsvektoren, mit welchen der zur Projektion benötigte (zweidimensionale) Unterraum aufspannt werden soll. Hierfür betrachten wir den bereits eingeführten dreidimensionalen Raum, in welchen bereits die relativen Spaltenanteile berücksichtigt sind. Dieser Raum wird bestimmt durch den Schwerpunkt und die projizierten Punkte (Vektoren), die sozusagen die äußere Abgrenzung des Raumes markieren. Dieses kann durch folgenden Ausdruck beschrieben werden:

$$(9) \quad \sum_{i=1}^7 (a_i^* - s^*) (a_i^* - s^*)^T$$

An dieser Stelle können die Massen (jetzt die relativen Anteile der Zeilen) mit in den Ausdruck einbezogen werden.

$$(10) \quad \sum_{i=1}^7 m_i (a_i^* - s^*) (a_i^* - s^*)^T$$

In dem durch Ausdruck (10) determinierten Raum soll der optimale Unterraum aufspannt werden. Da dieser ein Maximum an Varianz erklären soll, werden die Eigenwerte benötigt, um mit diesen die Einheitsvektoren bestimmen zu können. Kehren wir zu Gleichung (8) zurück. Dazu wurde gesagt, daß die Summe der u_i^2 maximal sein soll. Um dies zu erreichen, müssen wir Gleichung (5) über alle i maximieren, wobei die relativen Anteile der Zeilen (RPROZ - Zeilensummen in Prozent) berücksichtigt werden müssen. Betrachten wir daher folgenden Ausdruck.

$$(11) \quad e^T (\sum m_i (a_i^* - s^*) (a_i^* - s^*)^T) e, \text{ mit } e^T e = 1$$

Dieser Ausdruck wird maximal durch geeignete Wahl von e . Dies bedeutet, daß e der Eigenvektor zum größten Eigenwert von Ausdruck (10) ist. Dies soll in Matrixform dargestellt werden. Wir berechnen zuerst den Teilausdruck $(a_i^* - s^*)$ und nennen die

resultierende Matrix DIFFAIS (Zeilendarstellung minus Schwerpunkt). Aus dieser Matrix nehmen wir den ersten Zeilenvektor (AID und multiplizieren ihn derart mit sich selbst, daß die symmetrische Matrix AIII entsteht. Diese multiplizieren wir mit dem ersten Element des Spaltenvektors RPROZ (also mit 0,04663) und schreiben das Ergebnis in die Matrix AIIIM. Anschließend wird der gleiche Vorgang mit den nächsten Zeilenvektoren der Matrix DIFFAIS wiederholt, die sieben resultierenden symmetrischen Matrixen werden addiert und in die Matrix ERG (Zeilendarstellung, Matrix zur Berechnung der Eigenwerte) geschrieben. Aus dieser Matrix können nun die Eigenwerte (LAMBDA), die Eigenvektoren (EWERT) und die erklärte Varianz der Eigenwerte (EVAR) bestimmt werden.

ZEILENDARSTELLUNG - MINUS SCHWERPUNKT

DIFFAIS	COL1	COL2	COL3	COL4
ROW1	-0.123537	-0.39096	0.0612124	0.337003
ROW2	0.177049	-0.0140202	-0.0709178	-0.0365529
ROW3	-0.0835955	0.107595	0.0486976	-0.086997
ROW4	-0.0680889	-0.163599	0.0119874	0.174108
ROW5	0.0252259	0.108124	-0.0288333	-0.0688152
ROW6	-0.144863	0.00254863	-0.0713164	0.22065
ROW7	0.0982549	-0.0499192	0.110437	-0.198223

A11	COL1	COL2	COL3	COL4
ROW1	-0.123537	-0.39096	0.0612124	0.337003

AIII	COL1	COL2	COL3	COL4
ROW1	0.0152613	0.048298	-0.00756199	-0.0416323
ROW2	0.048298	0.15285	-0.0239316	-0.131755
ROW3	-0.00756199	-0.0239316	0.00374696	0.0206288
ROW4	-0.0416323	-0.131755	0.0206288	0.113571

AIIIM	COL1	COL2	COL3	COL4
ROW1	0.000711669	0.00225224	-0.000352632	-0.0019414
ROW2	0.00225224	0.00712771	-0.00111598	-0.00614401
ROW3	-0.000352632	-0.00111598	0.000174729	0.000961965
ROW4	-0.0019414	-0.00614401	0.000961965	0.00529607

ZEILENDARSTELLUNG MATRIX ZUR BERECHNUNG DER EIGENWERTE

ERG	COL1	COL2	COL3	COL4
ROW1	0.0140414	.0000615354	-0.00297255	-0.00762061
ROW2	.0000615354	0.0152714	-.000276124	-0.0121645
ROW3	-0.00297255	-.000276124	0.00370431	-0.00253238
ROW4	-0.00762061	-0.0121645	-0.00253238	0.0199485

EIGENWERTE DER MATRIX

ERKLAERTE VARIANZ DER EIGENWERTE

LAMBDA	COL1	EVAR	COL1
ROW1	0.0320625	ROW1	60.5345
ROW2	0.0148289	ROW2	27.9973
ROW3	0.00607419	ROW3	11.4682
ROW4	5.2044E-18	ROW4	9.8259E-15

EIGENVEKTOREN DER MATRIX

EWERT	COL1	COL2	COL3	COL4
ROW1	0.321077	-0.845852	-0.144301	0.400776
ROW2	0.555808	0.489454	-0.546501	0.39096
ROW3	0.0293574	0.211882	0.704056	0.677163
ROW4	-0.766241	0.00871613	-0.429906	0.477472

RANG DER MATRIX

R	COL1
ROW1	3

Aufgrund der Dreidimensionalität des Projektionsraumes ist der letzte Eigenvektor Null, (die in der Tabelle auftauchende Abweichung liegt im Rahmen der Rechengenauigkeit) der Rang der Matrix ist drei. Bestimmen lassen sich nun die

$$u_i = (a \cdot \overset{\cdot}{r} - s \cdot \overset{\cdot}{j}) \cdot e \cdot \overset{\cdot}{q} \cdot$$

womit die Projektion der "Studienfachvektoren" abgeschlossen wäre, die relative Lage der Zeilenvektoren im Projektionsraum ist durch die Matrix (UIZ) gekennzeichnet. Mit Hilfe der bisherigen Ergebnisse lassen sich die Kosinusse der Winkel (Phi) bestimmen (siehe Abbildung - grafische Darstellung der Projektion), sie entsprechen den Korrelationen der Variablenausprägungen mit den Achsen (CORRZZ). Der Anteil erklärter Varianz der Zeilenvektoren (CORR) ergibt sich aus dem Quadrat der Korrelationen, der Anteil, mit welchen die Zeilenvektoren die Achsen determinieren (SCHWERZZ) - in der Summe hier normiert auf eins - aus dem Quadrat der Entfernung zum Schwerpunkt multipliziert mit der relativen Häufigkeit der Zeilensummen. Ferner läßt sich der Anteil berechnen, mit welchen die Zeilenvektoren das gesamte Modell determinieren (INERTZ), indem die Anteile der Zeilenvektoren an den Achsen mit der Erklärungskraft der Achsen gewichtet werden.

KOORDINATEN DER ZEILEN

UIZ	COL1	COL2	COL3
THELOG	-0.513392	-0.0709558	0.129704
STATSW	0.0749803	-0.171964	-0.0521021
MED-PH	0.101052	0.132932	0.0249483
NATURW	-0.245849	-0.0184238	0.0328221
GEISTW	0.120078	0.0248753	-0.0534462
ARCHIT	-0.21626	0.110593	-0.125558
INGENE	0.158931	-0.0858702	0.176074

KORRELATION DER ZEILEN MIT DEN **ACHSEN**

CORRZZ	COL1	COL2	COL3
THELOG	-0.960948	-0.132812	0.242774
STATSW	0.385105	-0.883224	-0.267601
MED-PH	0.598529	0.787355	0.147768
NATURW	-0.988482	-0.0740764	0.131968
GEISTW	0.897656	0.185958	-0.399542
ARCHIT	-0.790916	0.404463	-0.459197
INGENE	0.630029	-0.340404	0.697988

ANTEIL ERKLAERTER VARIANZ DER ZEILEN

CORR	COL1	COL2	COL3
THELOG	0.923421	0.0176391	0.0589394
STATSW	0.148306	0.780084	0.0716102
MED-PH	0.358237	0.619928	0.0218354
NATURW	0.977097	0.00548731	0.0174155
GEISTW	0.805786	0.0345803	0.159634
ARCHIT	0.625547	0.16359	0.210862
INGENE	0.396937	0.115875	0.487188

ANTEIL DER ZEILEN AN DEN **ACHSEN**

SCHWERZZ	COL1	COL2	COL3
THELOG	0.383342	0.0158325	0.129152
STATSW	0.0431553	0.490799	0.109991
MED-PH	0.0973615	0.364288	0.0313247
NATURW	0.219768	0.00266853	0.0206761
GEISTW	0.047767	0.00443224	0.0499507
ARCHIT	0.147378	0.0833333	0.262229
INGENE	0.0612284	0.0386463	0.396676

RELATIVE ANTEILE ALLER ZEILEN

INERTZ	COL1
THELOG	0.251298
STATSW	0.176148
MED-PH	0.164521
NATURW	0.136154
GEISTW	0.0358849
ARCHIT	0.142619
INGENE	0.0933758

Im zweiten Teil der Aufgabe wären die "Organisationszugehörigkeitsvektoren" zu betrachten. Es läßt sich zeigen, daß sich bei analogem Vorgehen (d.h. es werden relative Spaltenanteile betrachtet und der Schwerpunkt aus den relativen Anteilen der Zeilen bestimmt) genau die gleichen Eigenwerte ergeben wie die eben berechneten, da das Ausgangsmaterial in beiden Fällen identisch ist. Geändert hat sich nur der Ausgangspunkt der Betrachtung. Ferner lassen sich sowohl der Schwerpunkt als auch die neuen Einheitsvektoren aus den zuvor berechneten Werten ableiten. Beweisen läßt sich dieser Sachverhalt mit dem Theorem von Eckart & Young (1936), zum rechnerischen Vorgehen siehe Greenacre (1984) oder Blasius (1986). Es ist daher zulässig, beide Teilaufgaben (Projektion der Zeilenvektoren und Projektion der Spaltenvektoren) in einer grafischen Darstellung zu repräsentieren und gemeinsam zu interpretieren.

Im folgenden werden die wichtigsten Ergebnisse der Spaltendarstellung der Beispielaufgabe angeführt, sowie die grafische Darstellung der Lösung.

KOORDINATEN DER SPALTENDARSTELLUNG

UJS	COL1	COL2	COL3
NSDAP	0.143452	-0.257008	-0.0280615
SS	0.25456	0.152452	-0.108944
SA	0.0077629	0.0381028	0.0810324
KMTG	-0.287354	0.00222296	-0.070173

KORRELATIONEN DER SPALTEN MIT DEN AXEN

CORRSS	COL1	COL2	COL3
NSDAP	0.48518	-0.869248	-0.094909
SS	0.805347	0.48231	-0.344663
SA	0.08637	0.423932	0.901566
KMTG	-0.971425	0.00751492	-0.237226

ANTEIL ERKLÄRTER VARIANZ DER SPALTEN

CORR1	COL1	COL2	COL3
NSDAP	0.2354	0.755593	0.00900771
SS	0.648584	0.232623	0.118793
SA	0.00745978	0.179718	0.812822
KMTG	0.943667	0.000056474	0.0562764

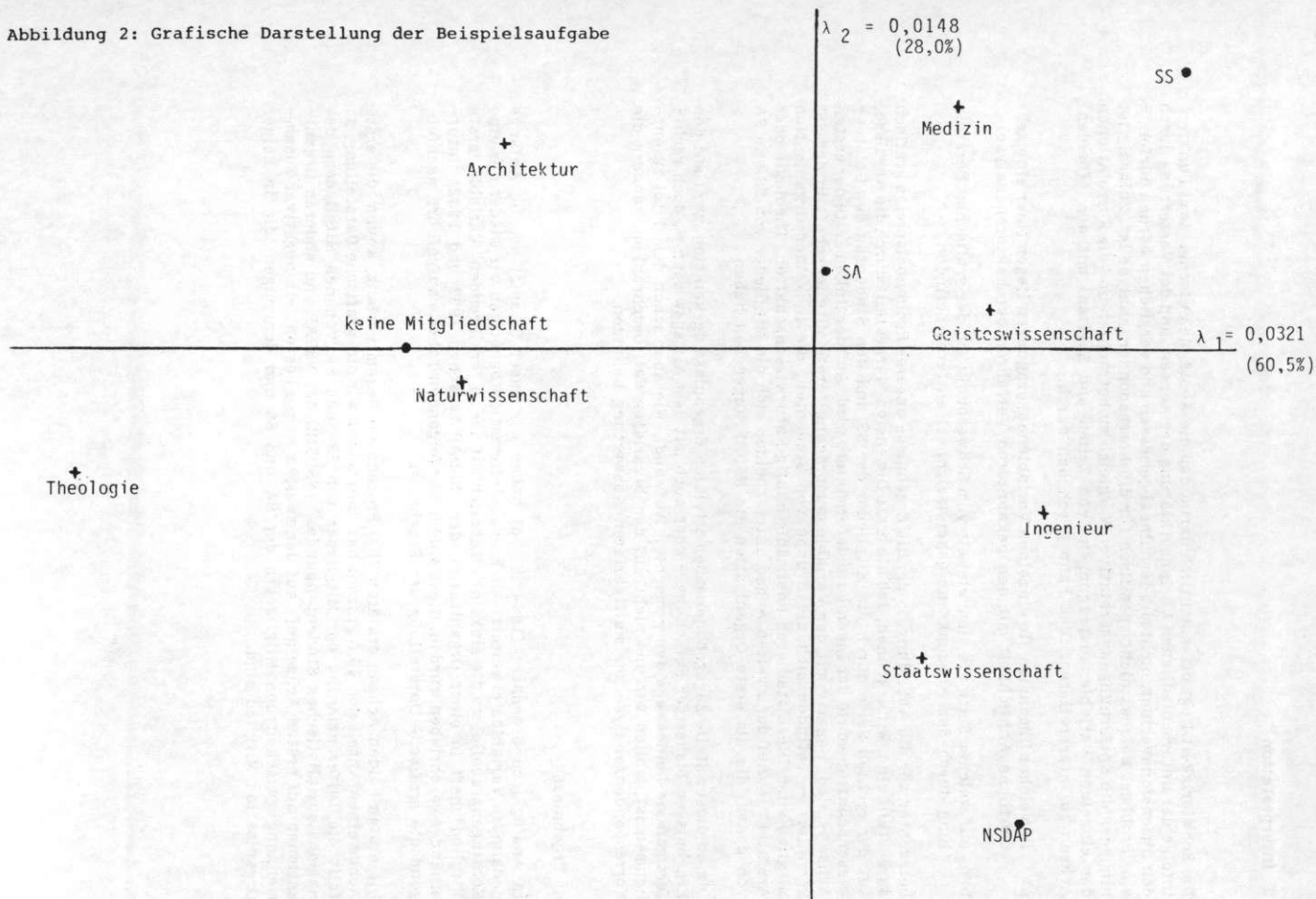
ANTEIL DER SPALTEN AN DEN AXEN

SCHWERS	COL1	COL2	COL3
NSDAP	0.103091	0.715465	0.0208227
SS	0.308922	0.239565	0.298663
SA	0.000861859	0.0448942	0.495695
KMTG	0.587125	0.0000759709	0.184819

RELATIVE ANTEILE ALLER SPALTEN

INERTS	COL1
NSDAP	0.265104
SS	0.288327
SA	0.069938
KMTG	0.37663

Abbildung 2: Grafische Darstellung der Beispielsaufgabe



8. Interpretation

Die Schwierigkeit liegt offensichtlich darin, daß die Abstände zwischen zwei Punkten, infolge der unterschiedlichen Varianzerklärung der Achsen und der Verzerrung durch den Dimensionsverlust, nicht mit dem Lineal gemessen und verglichen werden dürfen, - ein Problem, welches nicht spezifisch für die Korrespondenzanalyse ist. Zulässig ist hingegen die Entfernungen im relativen Vergleich zu betrachten und die Korrelationen der einzelnen Variablenausprägungen (von Zeilen und Spalten) mit den (beiden) Achsen zu interpretieren. Dies kann geschehen durch:

- a) in welcher Dimension die ~~einzelnen~~ Variablenausprägungen liegen (hier ~~eine~~ sehr deutliche Ähnlichkeit mit den bekannteren Verfahren der Faktorenanalyse)
- b) auf welcher Seite ~~sie sich wieweit~~ vom Schwerpunkt (in der grafischen Darstellung durch das Achsenkreuz gekennzeichnet) entfernt befinden.

Betrachten wir die Abbildung, in der die Ergebnisse unseres Rechenbeispiels grafisch dargestellt sind. Wir erkennen, daß die erste Dimension, dargestellt durch die Abszisse, auf der rechten Seite durch die Mitglieder der SS und den Studenten der Geisteswissenschaften sowie der Ingenieure determiniert wird. Auf der linken Seite der ersten Achse - also negativ mit der Variablenausprägung "erste Mitgliedschaft in der SS" korreliert - befinden sich die Gruppe von Studenten, die in keiner Organisation Mitglied sind. Diese studieren übermäßig häufig Naturwissenschaften, Theologie oder Architektur. Auf der zweiten Achse unterscheiden sich die Mitglieder von SS und SA von denen, die als erste Organisation die NSDAP angegeben haben.

Die an dieser Stelle zugegebenermaßen trivialen Ergebnisse decken sich zwar mit den Ergebnissen bisheriger Forschung - oder auch mit dem Alltagswissen -, doch rechtfertigen sie keineswegs den gemachten Aufwand. Wie aber schon dem Titel zu entnehmen ist, wollen wir uns nicht auf den bivariaten Fall beschränken, sondern die Korrespondenzanalyse zur multivariaten Auswertung benutzen.

4. Ergebnisse

In Analogie zu Greenacre (1984, S. 76) haben wir unser Beispiel um die Anfangs genannten Variablen erweitert. Als Eingabeinformation benötigen wir wieder nur eine Kontingenztafel, wo die einzelnen Tabellen mit der zu "erklärenden" Variable "erste Mitgliedschaft in einer Organisation der NSDAP zwischen 1933 und 1942" untereinander geschrieben wurden. Diese sind in der folgenden Tabelle aufgeführt, nachfolgend die grafische Darstellung der Ergebnisse.

Die ersten beiden Achsen erklären 75,0 Prozent der Gesamtvarianz, wovon die erste Achse einen Anteil von 49,0 Prozent hat. Betrachten wir die grafische Darstellung, so fällt die Differenzierung der Mitglieder von SS und SA von den Mitgliedern des nationalsozialistischen Studentenbundes (NSDStB), der NSDAP und anderen Organisationen auf ersten Achse auf. Auf der zweiten Achse besteht ein negativer Zusammenhang zwischen den Mitgliedern der SA und SS und denjenigen, die in keiner Organisation Mitglied sind.

AUSGANGSMATRIX (EINGABEDATEN)

A	NSDAP	SS	SA	NSDStB	SONST	KMTG
THELOG	2	0	9	2	0	7
STATSW	22	14	39	9	3	20
MED-PH	15	23	58	9	16	22
NATURW	6	4	21	2	3	14
GEISTW	7	8	18	5	4	8
ARCHIT	4	6	16	9	2	13
INGENE	6	4	16	7	6	4
S-FACH	7	4	12	3	3	1
MAENNL	74	64	195	50	37	92
WEIBLI	5	0	0	18	11	11
D33-38	31	32	151	26	29	73
D39-42	48	32	45	42	20	32
V-DEUT	66	45	163	52	36	69
V-AUSL	10	18	25	12	11	26
B-1904	3	6	5	1	3	18
1905-9	13	6	45	7	7	33
1910-4	35	44	125	33	26	36
B-1925	28	8	21	26	13	16
OS-ADL	2	5	11	2	3	4
BEA-HS	5	11	18	0	6	18
SON-HS	7	6	19	10	3	6
ALT-MS	14	14	51	11	8	32
NEU-MS	35	22	75	19	15	22
ARBEIT	9	4	17	5	2	7

Quelle: Arminger 1984: S.20-25

Sollen die Mitglieder der NSDAP beschrieben werden, so läßt sich hier z.B. die Gruppe der Ingenieure zuordnen, sowie jene Studenten, deren Väter Akademiker und nicht im Staatsdienst sind. (z.B. Ärzte, Journalisten, ...) Der NSDStB lassen sich insbesondere jüngere Studenten (Geburtskohorte 1915 bis 1925) zuordnen und jene Personengruppen, die ihr erstes Darlehen zwischen 1939 und 1942 erhielten.

Die grafische Darstellung soll nicht genauer interpretiert werden, wir wollten damit lediglich einen ersten Überblick über die Zusammenhänge geben. Eine exaktere Interpretationsmöglichkeit - und zwar des gesamten Projektionsraumes - ist mit Hilfe der numerischen Ergebnisse möglich, die in Tabelle 1 aufgeführt sind.

Die Spalten, in denen die quadrierten Korrelationen der Variablenausprägungen (von Zeilen und Spalten) mit den Achsen dargestellt sind (qKor), sind die wichtigsten Interpretationshilfen. Hier ist abzulesen, auf welcher Achse welche Variablenausprägungen positiv oder negativ (das Vorzeichen ist den Spalten "Lage" zu entnehmen) "lädt": Die Interpretation ist analog der Hauptkomponentenanalyse. Der Vorteil gegenüber dieser Methode besteht darin, daß sich neben den zu interpretierenden Variablen, auch die Untersuchungseinheiten (hier "Organisationszugehörigkeit") parallel interpretieren lassen. Die Summe der quadrierten Korrelationen der ersten vier (der fünf existierenden) Achsen ist in der Spalte "SqKor" dargestellt.

In der Spalte "Masse" befindet sich der relative Anteil von Zeilen und Spalten, getrennt auf jeweils tausend normiert. In dieser Spalte werden auch fehlende Werte berücksichtigt, in diesem Fall ist die Gesamtmasse der betreffenden Variable geringer

Abbildung 3: Grafische Darstellung der Ergebnisse

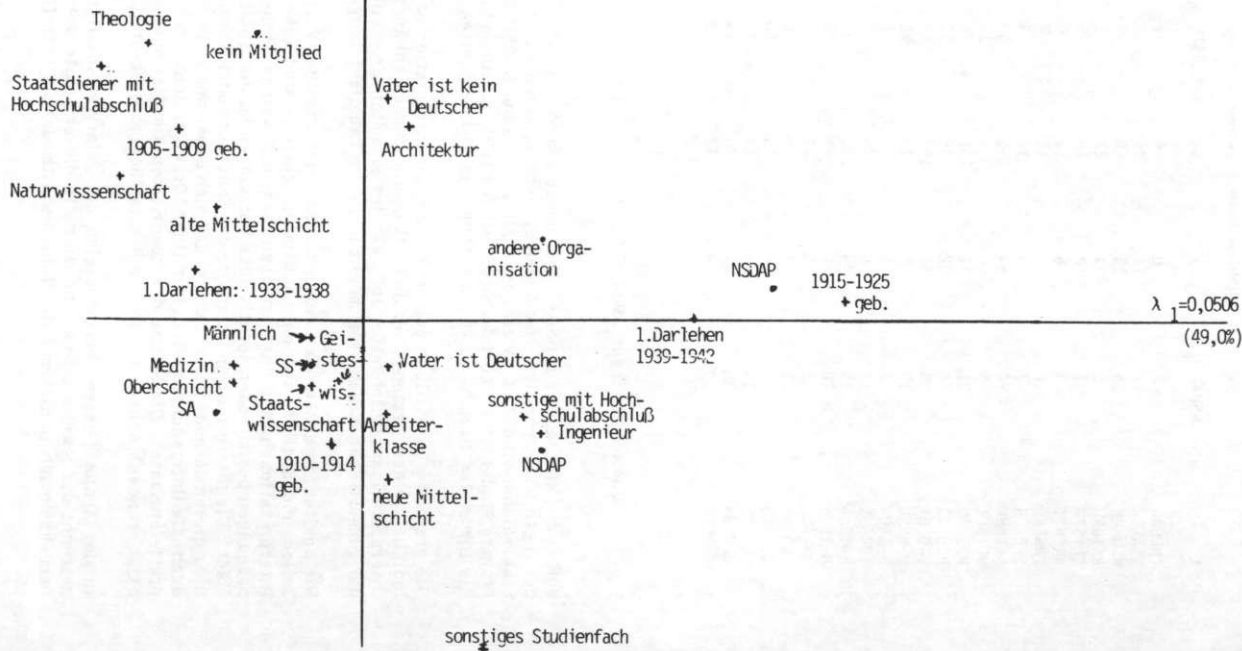


Tabelle 1: Numerische Darstellung der Ergebnisse

Variable	Gesamt SqKor Mass Trg.			1. Achse Lage qKor Trg.			2. Achse Lage qKor Trg.			3. Achse Lage qKor Trg.			4. Achse Lage qKor Trg.		
THELOG	963	6	24	-274	193	9	301	232	21	-300	232	49	346	306	69
STATSW	991	33	24	-60	51	2	-81	91	8	125	212	46	217	637	145
MED-PH	953	45	29	-144	319	19	-69	75	8	60	54	14	-182	505	137
NATURW	970	16	20	-288	638	26	155	183	14	-60	29	5	126	120	23
GEISTW	954	16	3	-27	43	0	-61	206	2	101	543	14	-54	162	4
ARCHIT	427	16	19	42	14	1	222	385	29	-52	22	4	28	6	1
INGENE	952	13	14	203	374	11	-132	161	9	-145	195	25	-155	222	30
S-FACH	894	9	17	117	72	3	-390	799	53	66	23	4	9	0	0
MAENNL	985	160	8	-55	619	10	-31	210	6	19	69	5	21	87	6
WEIBLI	998	14	211	1020	670	289	579	216	176	-333	72	136	-249	40	81
D33-38	995	107	61	-199	676	84	48	39	9	-122	256	140	-37	24	14
D39-42	997	68	119	386	833	202	-14	1	1	163	148	157	52	15	17
V-DEUT	979	135	9	30	130	2	-56	484	17	-45	308	24	20	57	5
V-AUSL	984	32	35	33	9	1	247	531	73	174	263	84	-143	181	61
B-1904	998	11	87	-243	75	13	784	771	257	342	147	114	60	5	4
1905-9	978	35	48	-230	371	37	218	329	61	-130	118	51	152	160	74
1910-4	943	93	38	-41	43	3	-150	547	79	-7	2	1	-120	351	125
B-1925	982	35	113	557	931	214	16	1	0	-10	0	0	129	50	54
OS-ADL	996	8	9	-152	226	4	-65	41	1	76	55	4	-264	674	55
BEA-HS	967	18	49	-321	373	37	287	297	56	260	244	107	-121	53	25
SON-HS	609	16	14	181	354	10	-124	167	9	-89	88	11	-4	0	0
ALT-HS	966	41	18	-167	603	23	118	295	21	-34	26	4	44	42	7
NEU-HS	927	59	24	18	8	0	-192	872	81	8	2	0	44	45	10
ARBEIT	999	14	7	24	10	0	-120	258	7	-17	6	0	202	725	52
NSDAP	958	142	135	196	393	108	-131	179	93	90	83	100	172	303	389
SS	961	119	94	-58	42	8	-64	51	19	239	693	587	-119	175	157
SA	998	361	191	-186	636	248	-114	243	179	-80	119	204	-3	0	0
NSDSTB	966	113	273	481	923	513	33	4	5	-97	38	94	-14	1	2
SONST	838	83	92	193	330	62	77	52	18	-40	15	12	-223	441	385
KMTG	1000	183	216	-129	139	61	318	827	687	14	2	3	63	32	67

als die Gesamtmasse einer Variable ohne fehlende Werte. So ist z.B. die Gesamtmasse der Variable "Geschlecht" 174, während sie bei der Variable "Nationalität des Vaters" nur 167 ist, (es sind die einzelnen Werte jeweils zu addieren).

Anhand der "Massenträgheit" (Trg. in der Tabelle) wird ersichtlich, wie stark das Modell von den einzelnen Variablenausprägungen (für Zeilen und Spalten getrennt wiederum auf tausend normiert) determiniert wird. Es ist ähnlich einer Apothekerswaage, wo ein größeres Gewicht (hier als Anzahl von Personen) durch eine entsprechend größere Entfernung vom Auflagepunkt ausgeglichen werden kann. Die gleiche Interpretation gilt auch für die jeweiligen Achsen.

Die Spalten "Lage" geben die relative Lage der Variablenausprägungen von Zeilen und Spalten auf den einzelnen Achsen an. Hier ist ablesbar, auf welcher Seite der Achse sich die einzelnen Punkte wie weit vom Schwerpunkt (dargestellt durch das Achsenkreuz) entfernt befinden.

Auf der ersten Achse, die 49,0 Prozent der Varianz erklärt, "laden" (als "Schwellenwert" für den Anteil erklärter Varianz durch die einzelnen Achsen wählen wir 35 Prozent) im positiven Bereich die Frauen, diejenigen die zwischen 1915 und 1925 geboren wurden, die ihr erstes Darlehen zwischen 1939 und 1942 erhielten, sowie

(allerdings etwas schwächer) Studenten des Ingenieurwesens und jene Personengruppe, deren Väter einen Hochschulabschluß hatten, aber nicht im öffentlichen Dienst beschäftigt waren. Dies korrespondiert auf der Ebene der zu erklärenden Variable "Organisationszugehörigkeit" mit den Mitgliedern des NSDStB und (schwächer "ladend") der NSDAP.

Im negativen Bereich der ersten Achse, also negativ korreliert mit der Mitgliedern von NSDAP und NSDStB befinden sich die Mitglieder der SA. Sie ließen sich als Studenten der Naturwissenschaften beschreiben, deren Mitglieder relativ häufig zwischen 1905 und 1909 geboren sind, und die ihr erstes Darlehen zwischen 1933 und 1938 erhielten. Der Vater ist oft Angehöriger der alten Mittelschicht (Farmer, Handwerker) bzw. (aber schwächer "ladend") Staatsbediensteter mit Hochschulabschluß.

Auf der zweiten Achse, die entsprechend dem Modell unkorreliert mit der ersten ist und die 26,0 Prozent der Varianz erklärt, "laden" im positiven Bereich Studenten der Architektur. Sie sind relativ häufig vor 1904 geboren, der Vater ist übermäßig häufig Ausländer. Auf der Ebene der "Mitgliedschaft bei einer Organisation" befindet sich jene Gruppe von Personen, die in keiner der aufgeführten Organisationen Mitglied sind.

Mit dieser Gruppe der "Nichtmitglieder" korreliert keine der berücksichtigten Organisationen negativ. Insofern läßt sich nur eine Aussage darüber machen, welche "Eigenschaften" bei den Personen nicht vorhanden sind, die in keiner Organisation Mitglied sind. Dies sind Personengruppen, deren Väter deutscher Staatsangehörigkeit und relativ häufig der neuen Mittelschicht (z.B. Grundschullehrer) zuzurechnen sind. Ebenso "lädt" im negativen Bereich der ersten Achse die Gruppe der zwischen 1910 und 1914 Geborenen, sowie die Studenten anderer Fachrichtungen.

Auf der dritten Achse, die noch 11,2 Prozent der Gesamtvarianz erklärt, laden im positiven Bereich die Mitglieder der SS, von denen relativ viele Studenten der Geisteswissenschaften sind. Zu dieser Gruppierung gibt es keine negativ korrelierende Gruppe von Personen, (sowohl auf der Ebene der "abhängigen" als auch der "unabhängigen" Variablenausprägungen).

Auf der vierten Achse, die 10,5 Prozent der Varianz erklärt, "laden" die Mitglieder "sonstiger nationalsozialistischer Organisationen". Sie lassen sich beschreiben als Studenten der medizinischen Fächer, als Gruppe von Personen, die (schwächer "ladend") zwischen 1910 und 1914 geboren sind und deren Väter übermäßig oft der Oberschicht (Adlige, Topmanager, Offiziere, ...) zuzurechnen sind. Dazu konträr befinden sich die Studenten der Staatswissenschaften (Jura, Ökonomie, Soziologie und Politologie) und die Gruppe von Studenten, deren Väter übermäßig oft der Arbeiterklasse zuzuordnen sind. Eine Mitgliedschaft in einer bestimmten Organisation ist auf dem positiven Abschnitt der vierten Achse nicht vorhanden.

Auf der Ebene der erklärenden Variablen wird der Projektionsraum insbesondere durch die weiblichen Studenten (Trg.=211), durch diejenigen, die erstes Darlehen zwischen 1939 und 1942 bekamen (Trg.=119) und die zwischen 1915 und 1925 geboren (Trg.=113) sind determiniert. Hierbei ist zu berücksichtigen, daß bei der Determination des Gesamtmodelles eine Gewichtung über die Erklärungskraft der einzelnen Achsen erfolgt. So ist der starke "Einfluß" der genannten Variablenausprägungen auf ihre hohe Erklärungskraft bei der ersten Achse zurückzuführen, die sie gemeinsam zu 70,5 Prozent determinieren. Dies korrespondiert auf der Ebene der "zu erklärenden" Variablen mit der Mitgliedschaft im NSDStB, die hier die erste Achse zu 51,3 Prozent (Trg.=513) determiniert. Diese genannten Gruppen bestimmen somit am stärksten den Interpretationsraum, d.h. sie sind die "wichtigsten" Variablen zur Beschreibung der aufgezeigten Zusammenhänge.

5. Fazit

An dieser Stelle könnte die inhaltliche Interpretation der gezeigten Zusammenhänge erfolgen, doch war dies nicht das Ziel dieser Arbeit. Vielmehr sollte und konnte gezeigt werden, daß die Korrespondenzanalyse ein sehr gutes Instrument zur multivariaten Auswertung qualitativer Daten ist. Es können Hypothesen getestet werden, für die bislang ein geeignetes Verfahren fehlte. Die Korrespondenzanalyse sollte auch Grund sein, zu prüfen, ob in Anwendungsbeispielen wie dem hier vorgestellten, nicht völlig auf Verfahren wie Faktoren-, Cluster- oder Diskriminanzanalyse verzichtet werden kann oder muß.

6. Literatur

- ARMINGER, G. (1984), Involment of German Students in NS Organisations Based on the Archive of the Reichsstudentenwerk. *Historical Social Research* 30 (1984), 3-34.
- BENZECRI, J.-P. (1963), *Course de Linguistique Mathématique*. Rennes: Université de Rennes.
- BLASIUS, J. (1986), Korrespondenzanalyse. Unveröffentlichtes Arbeitspapier. Köln: Universität zu Köln.
- BLASIUS, J./ROHLINGER, H. (1987), KORRES - Ein Programm zur Korrespondenzanalyse. Vortragspapier zur Statistik-Software Konferenz in Heidelberg vom 23.3 bis 26.3 1987. Köln: Universität zu Köln.
- DANGSCHAT, J./ BLASIUS, J. (1987), Social and Spatial Disparities in Warsaw 1978: An Application of Correspondence Analysis to a 'Socialist' City. *Urban Studies* 24, 173-191.
- ESCOFIER-CORDIER, B. (1965), *L' Analyse des Correspondances*. Unveröffentlichte Dissertation: Université de Rennes.
- GREENACRE, M. (1981), Practical Correspondence Analysis. In: Barnett, V. 1981, 119-146.
- GREENACRE, M. (1984), *Theory and Applications of Correspondence Analysis*. London: Academic.
- GUTTMAN, L., (1941), The Quantification of a Class of Attributes: A Theory and Method of Scale Construction. In: Horst, P. 1941, 319-348.
- GUTTMAN, L. (1950), The Principal of Scale Analysis. In: Stouffer, S.A. et al. 1950.
- HAYASHI, C. (1950), On the Quantification of Qualitative Data from the Mathematico - Statistical Point of View. *Annals of the Institute of Statistical Mathematics* 2, 35-47.
- HAYASHI, C. (1952), On the Prediction of Phenomena from Qualitative Data and the Quantification of Qualitative Data from the Mathematico-Statistical Point of View. *Annals of the Institute of Statistical Mathematics* 3, 69-98.
- HAYASHI, C. (1954), Multidimensional Quantification - with the Applications to Analysis of Social Phenomena. *Annals of the Institute of Statistical Mathematics* 5, 231-245.
- HILL, O. (1973), Reciprocal Averaging: An Eigenvector Method of Ordination. *Journal of Ecology* 61, 237 -251.
- HIRSCHFELD, H.O. (1935), A Connection between Correlation and Contingency. *Proceedings of the Cambridge Philosophical Society* 31, 520-524.
- NISHISATO, S. (1978), Optimal Scaling of Paired Comparison and Rank Order Data. An Alternative to Guttman's Formulation. *Psychometrika* 43, 263-271.
- Nishisato, S. (1980), *Analysis of Categorical Data: Dual Scaling and its Applications*. Toronto: University of Toronto Press.
- SAS 1985; *User's Guide: Statistics*. Gary, N.C.: SAS Institute Inc.